Figure 1. Process Flow Chart

LaSSI Database Construction

/ S100

Create chemical descriptors
from connection tables

↓

/ S110

Create index of unique
descriptors

↓

/ S120

Create descriptor-molecule
matrix

↓

/ S130

Perform singular value
decomposition of matrix

Query Handling

/ S200

Allow user to specify query
compound(s)

↓

/ S210

Create chemical descriptors
for query compound(s)

↓

/ S220

Transform query into multi-
dimensional space using
SVD matrices

↓

/ S230

Calculate similarity
between query and database
compounds

↓

/ S240

Sort compounds by
similarity to query
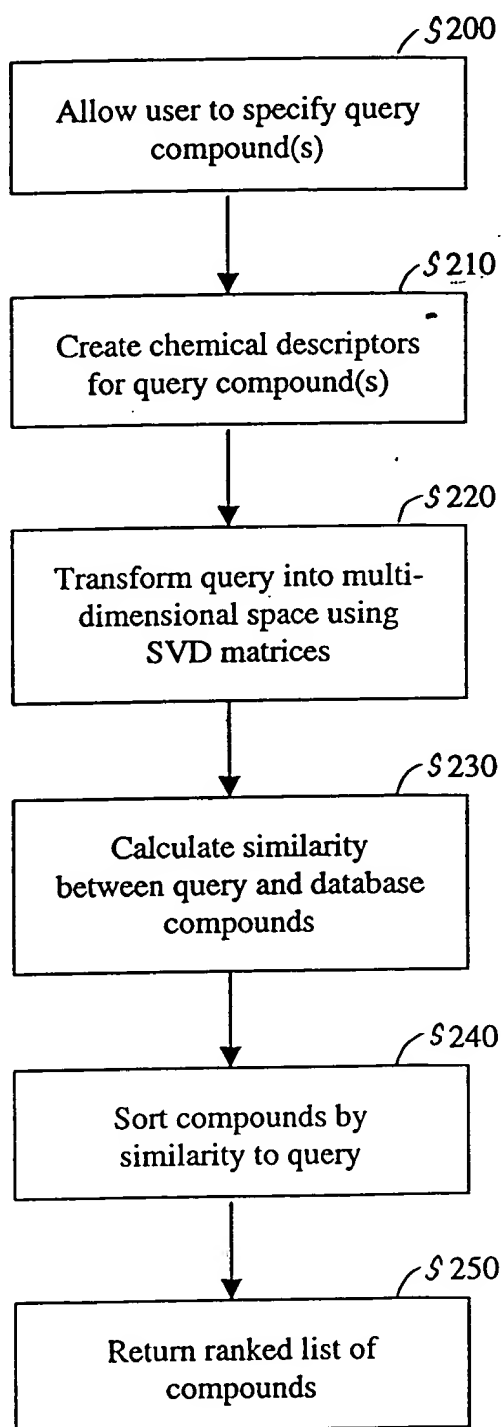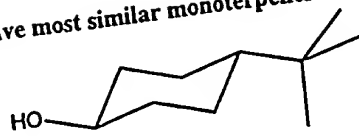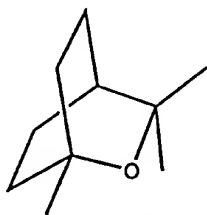
↓

/ S250

Return ranked list of
compounds

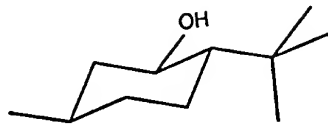Figure 2. Probe and its twelve most similar monoterpenes selected using 2 singular values

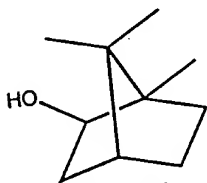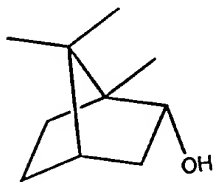4-*t*-butylcyclohexanol
Probe molecule

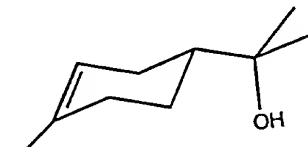terpin
0.531

eucalyptol
0.435
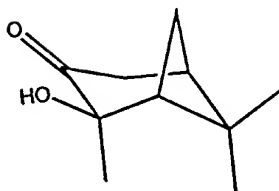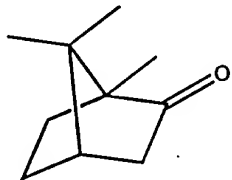
menthol
0.389

isoborneol
0.389

borneol
0.389
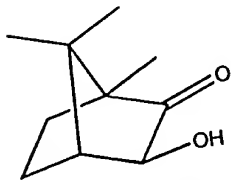
α-terpinol
0.361

6 Most similar by Tanimoto similarity
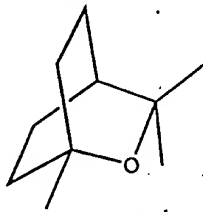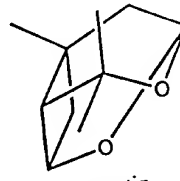
oxipinocamphone
1.0000

camphor
0.9998
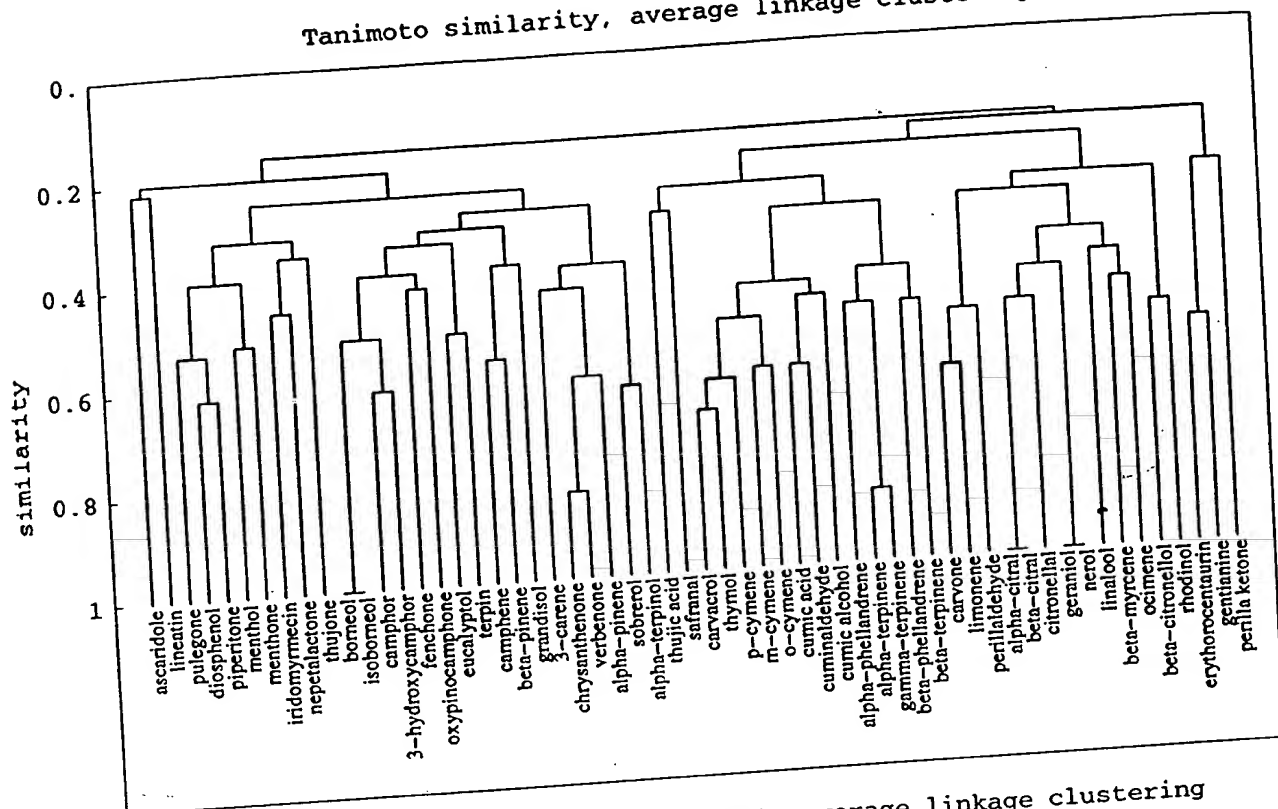
terpin
0.9997

3-hydroxycamphor
0.9997

eucalyptol
0.9995

lineatin
0.9992

6 Most similar by LaSSI similarity
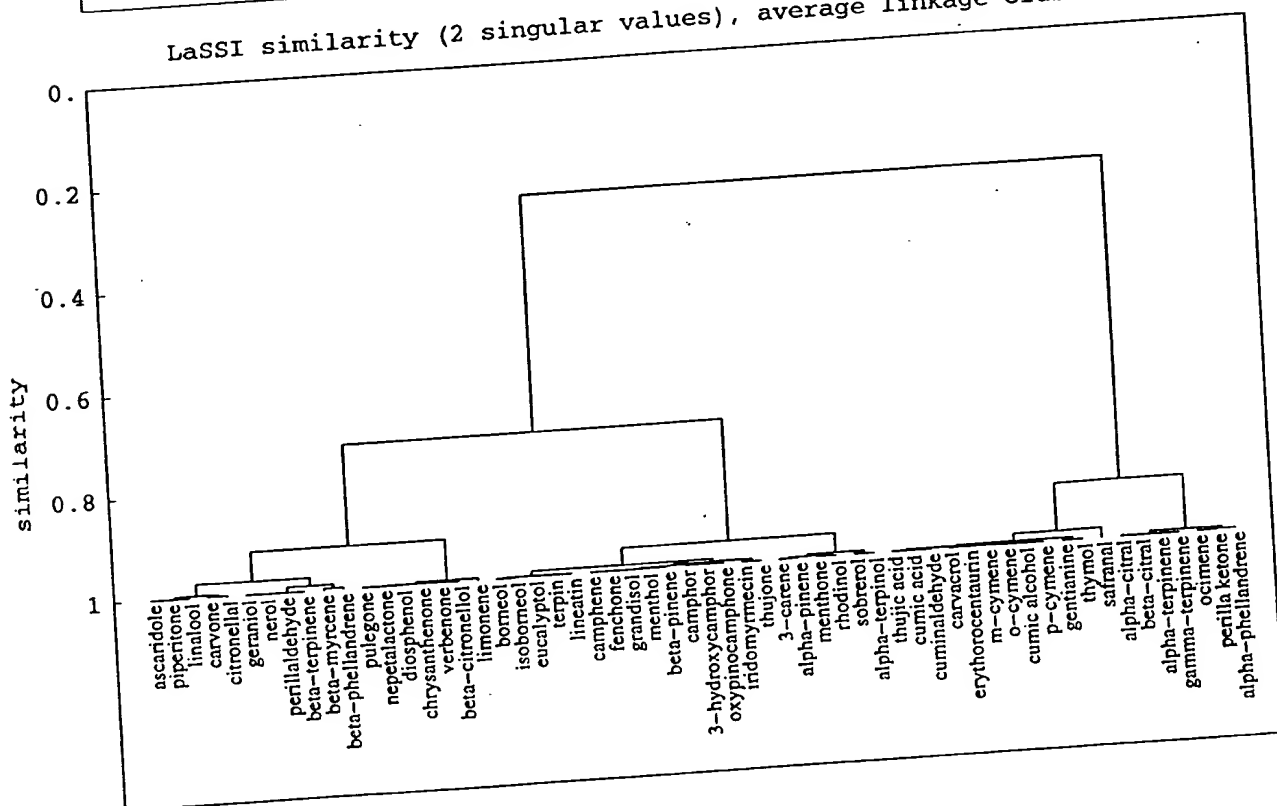
Figure 3. D ndrograms Showing Similarities For Tanimoto ans LaSSI

Figure 4. Two–dimensional Plot of Example Database Compounds and Probe Compound

S300

NO

Has a user input a query compound probe?

↓ YES

S310

Create chemical descriptors for the query compound probe.

S320

YES

Has the query compound been modified in view of the ~~created chemical descriptors~~ ranked compounds

✳

↓ NO

S330

Transform the modified query probe into multi-dimensional space using singular value decomposition matrices.

↓

340

Calculate similarity between query probe and compounds database.

↓

S350

Rank compounds by similarity to query probe.
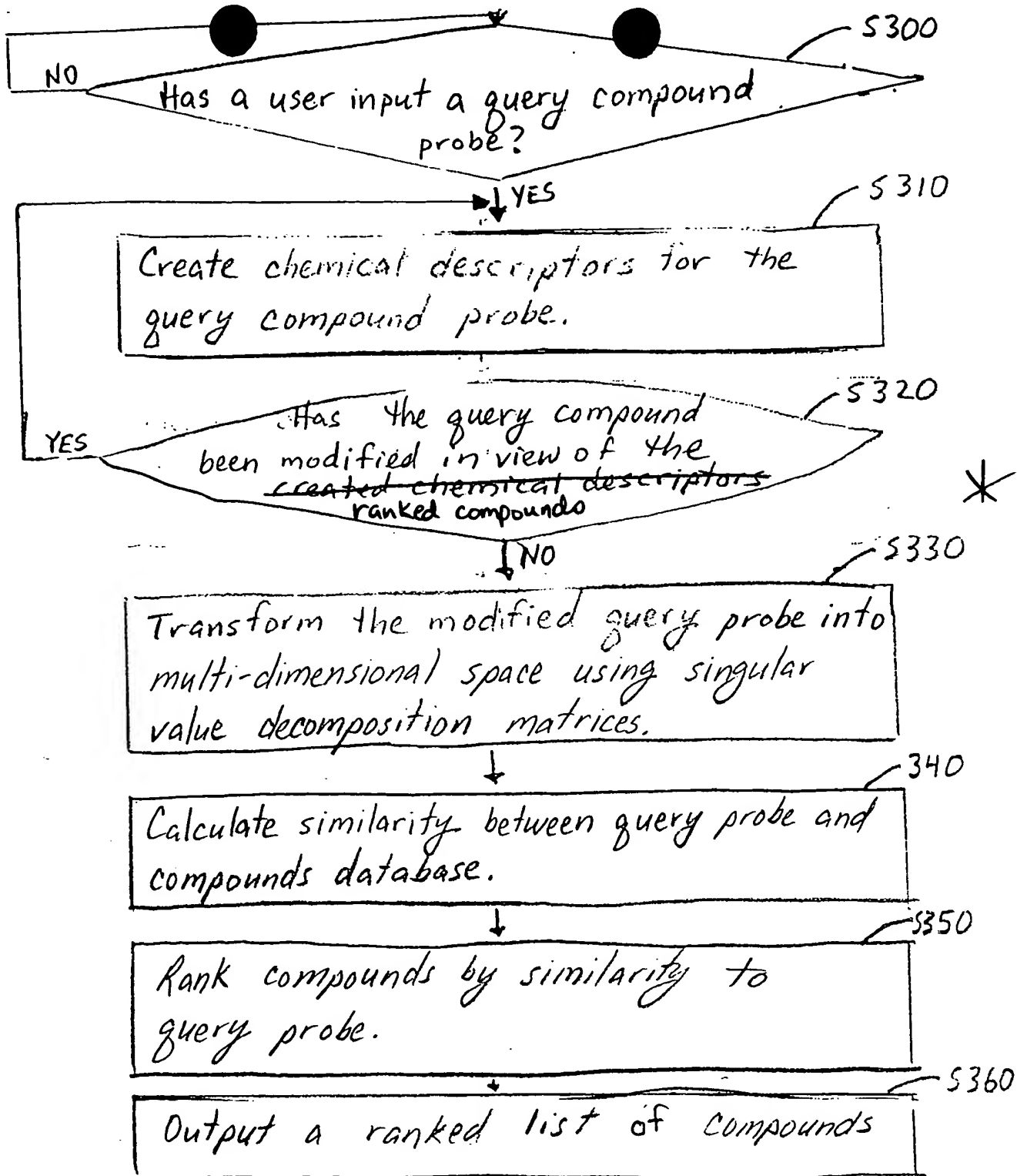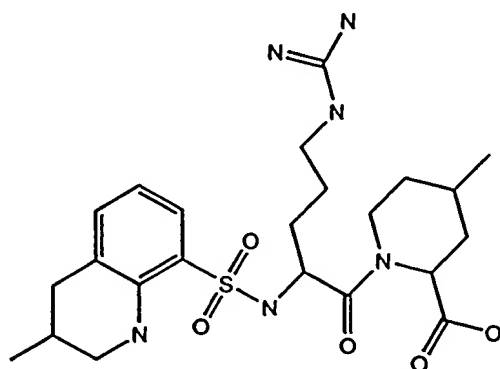
S360

Output a ranked list of compounds

Figure 5

SURESH,
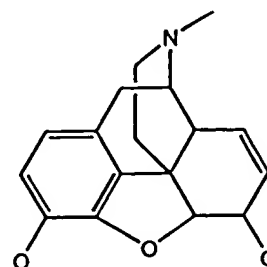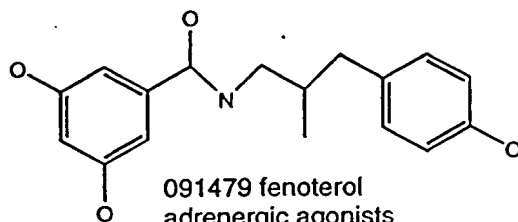S320 is really not right. We need to talk about this more.   Richard

Figure 6a.

090744 argatroban
thrombin inhibitors
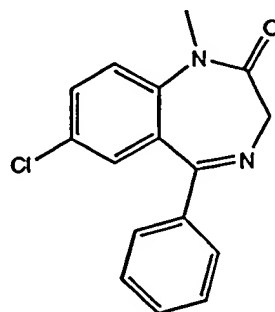
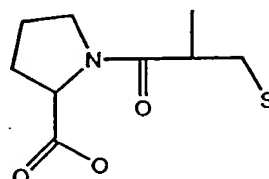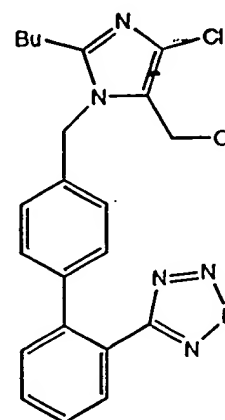091323 diazepam
anxiolytics

091342 morphine
opioid analgesics

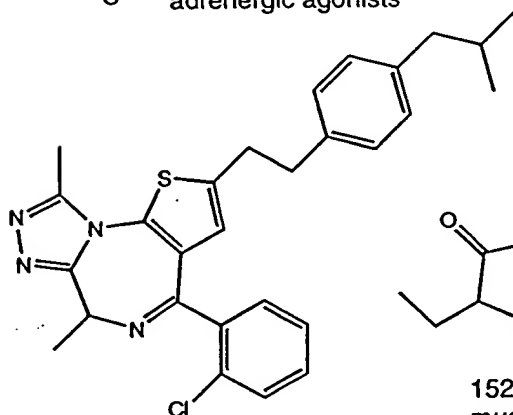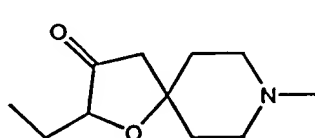091479 fenoterol
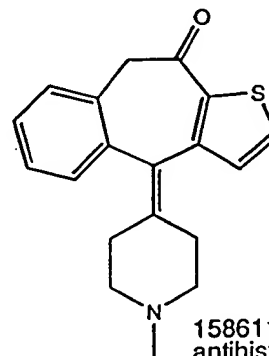adrenergic agonists

115230 captopril
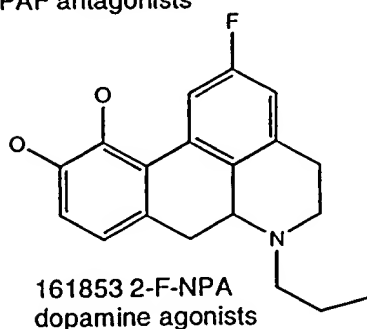ACE inhibitors

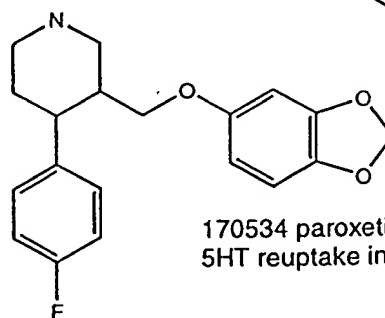140603 losartan
AII blockers

144822 israpafant
PAF antagonists

152580 YM-954
muscarinic agonists

158611 ketotifen
antihistamines

161853 2-F-NPA
dopamine agonists

170534 paroxetine
5HT reuptake inhibitors

170958
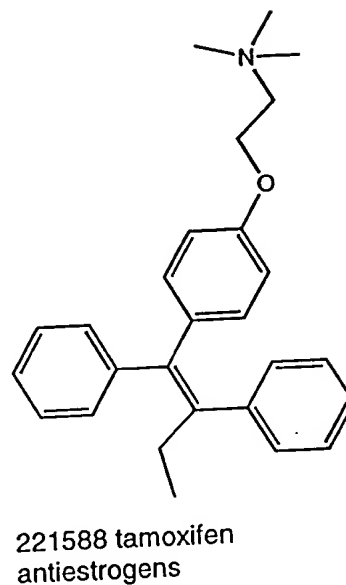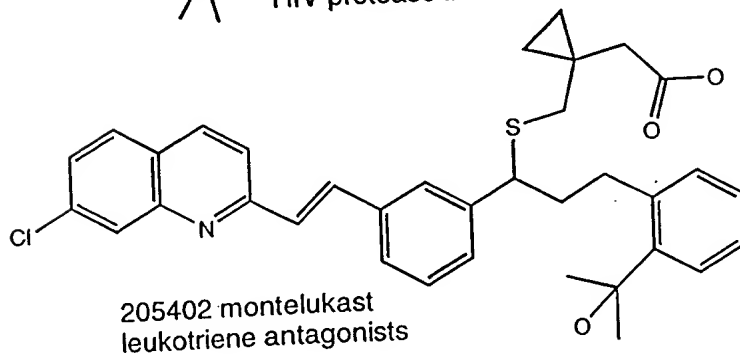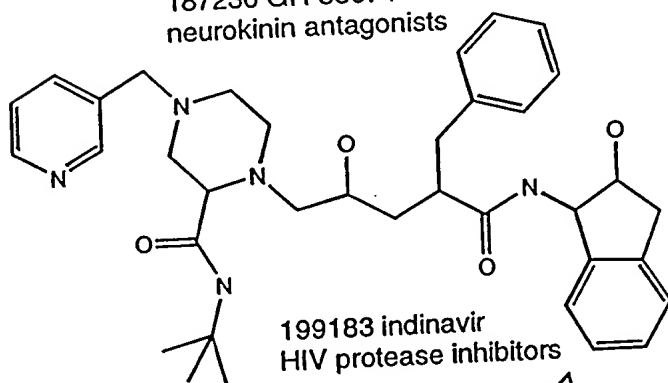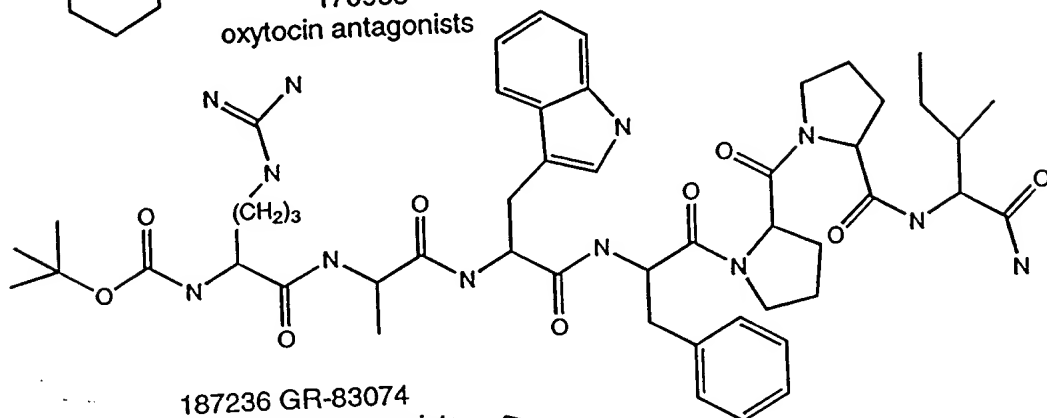oxytocin antagonists

187236 GR-83074
neurokinin antagonists

199183 indinavir
HIV protease inhibitors

205402 montelukast
leukotriene antagonists

221588 tamoxifen
antiestrogens
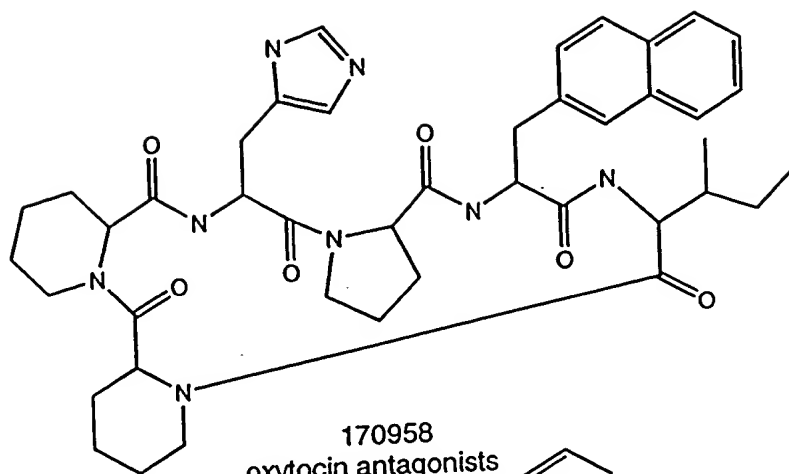
Figure 6b

Figure 7   Probes used for peptide -> non-peptide tests.

159880 F-DPDPE
opioid analgesics

187236 GR-83074
neurokinin antagonists

170958
oxytocin antagonists

188541 G-4120
gpIIb/IIIa receptor antagonists

174556 BQ-123
endothelin antagonists

cycAII [Sar^1,Hcy^{3,5},Ile^8]AII
AII blockers
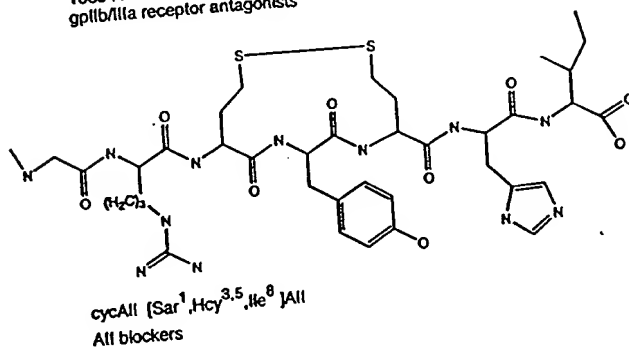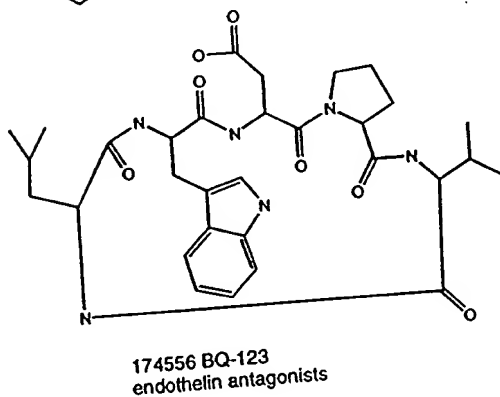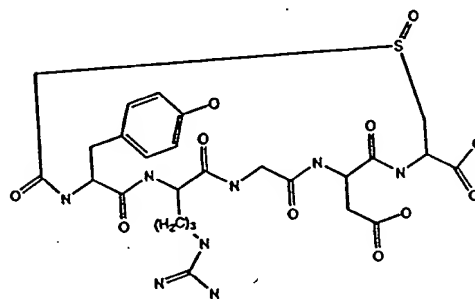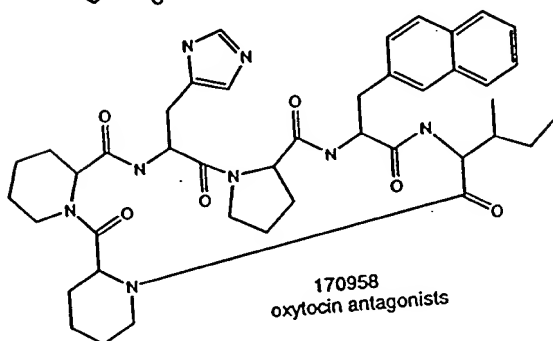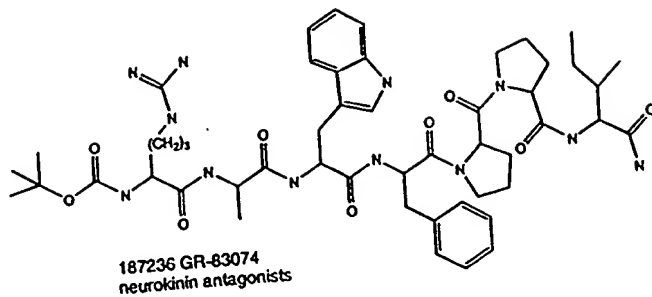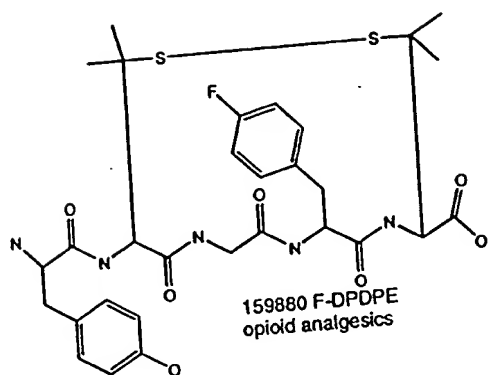
Figure 8. The initial enhancement for LaSSI APTT vs the number of singular values shown for three examples.

Initial Enhancement vs Singular Values

199183 as probe

HIV-1 protease inhibitors as actives

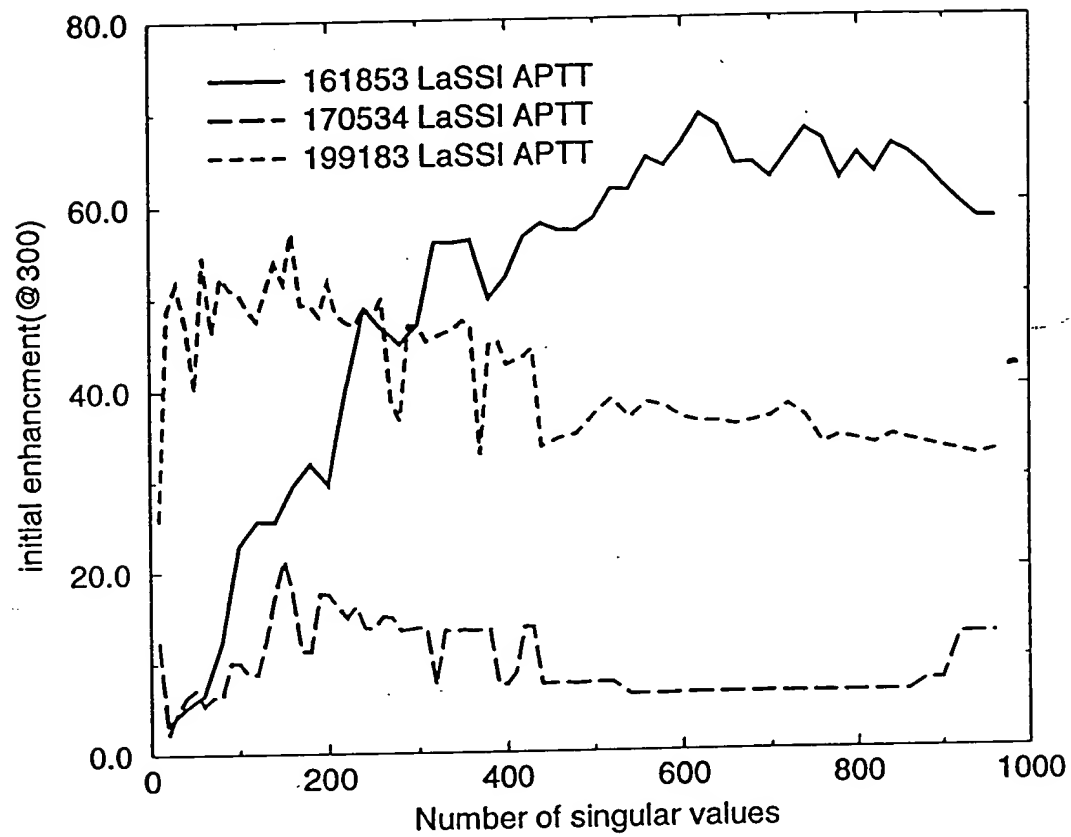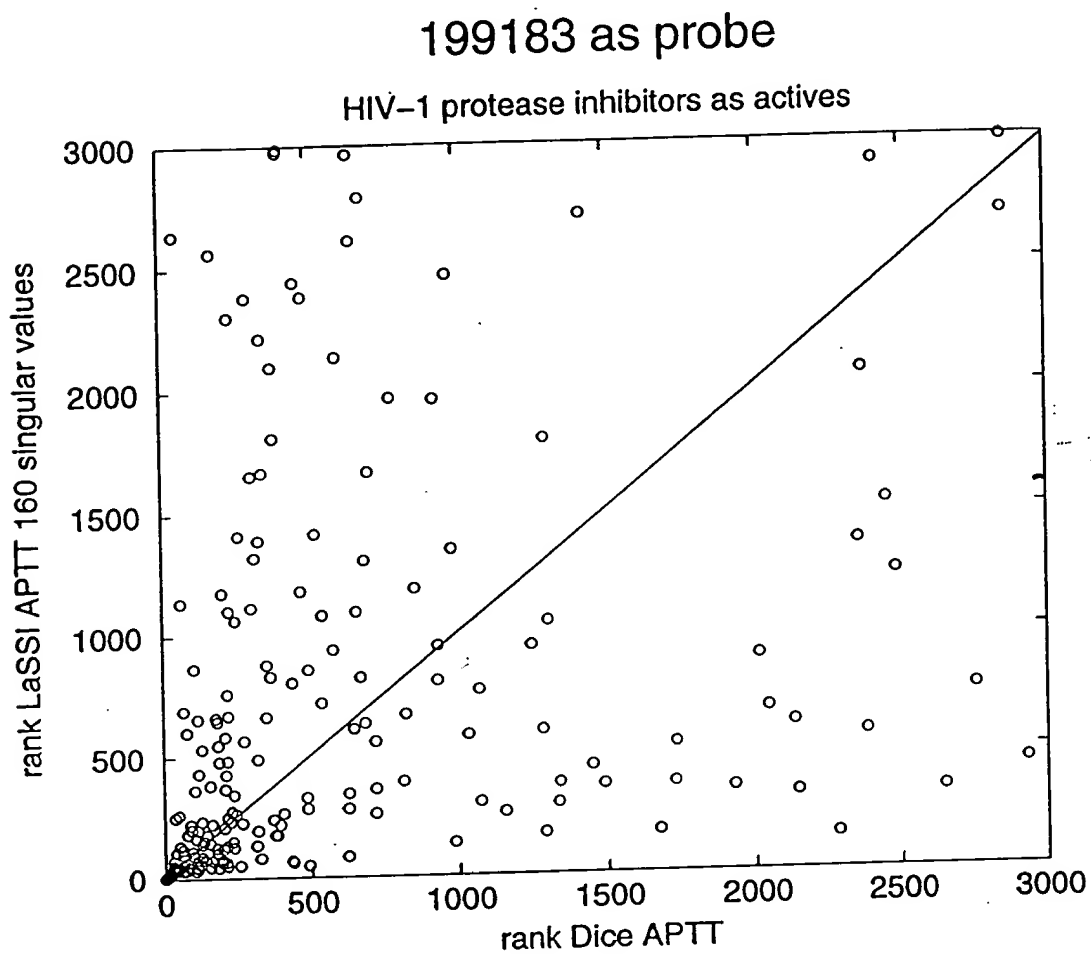Figure **1**. Selected compounds that have extremely different ranks in Dice APTT vs LaSSI APTT. The examples are 161853 with 800 singular values, 170534 with 150 singular values. The ranks in tw types of search are indicated.
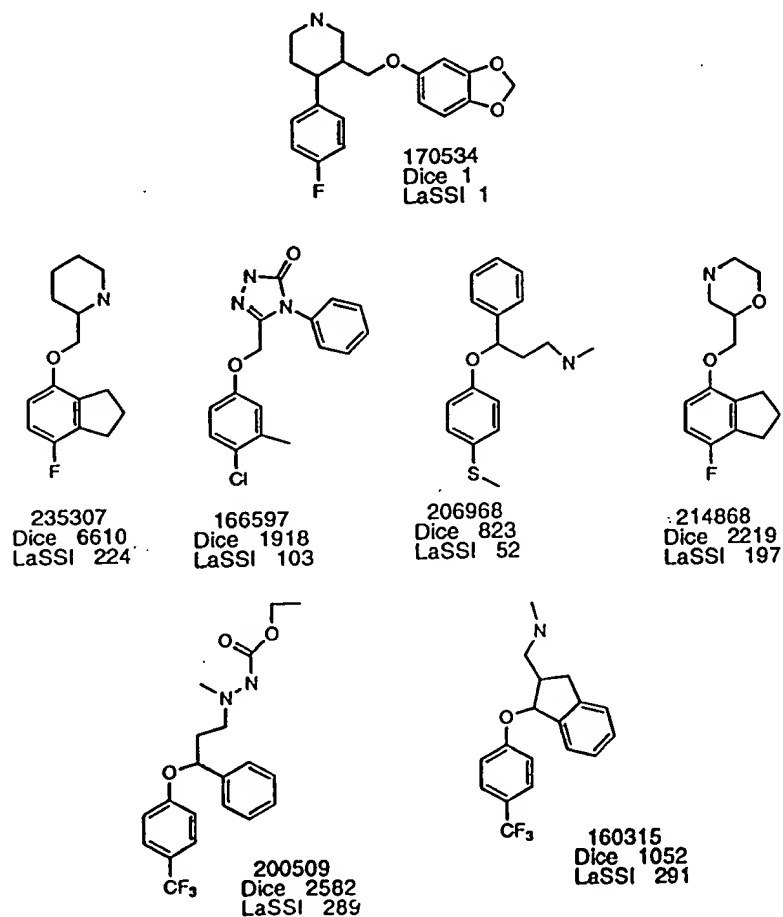
161853
Dice 1
LaSSI 1

174007
Dice 265
LaSSI 31

222386
Dice 692
LaSSI 90

179378
Dice 936
LaSSI 127

172510
Dice 483
LaSSI 82

177101
Dice 735
LaSSI 160

170534
Dice 1
LaSSI 1

235307
Dice 6610
LaSSI 224

166597
Dice 1918
LaSSI 103

206968
Dice 823
LaSSI 52

214868
Dice 2219
LaSSI 197

200509
Dice 2582
LaSSI 289

160315
Dice 1052
LaSSI 291
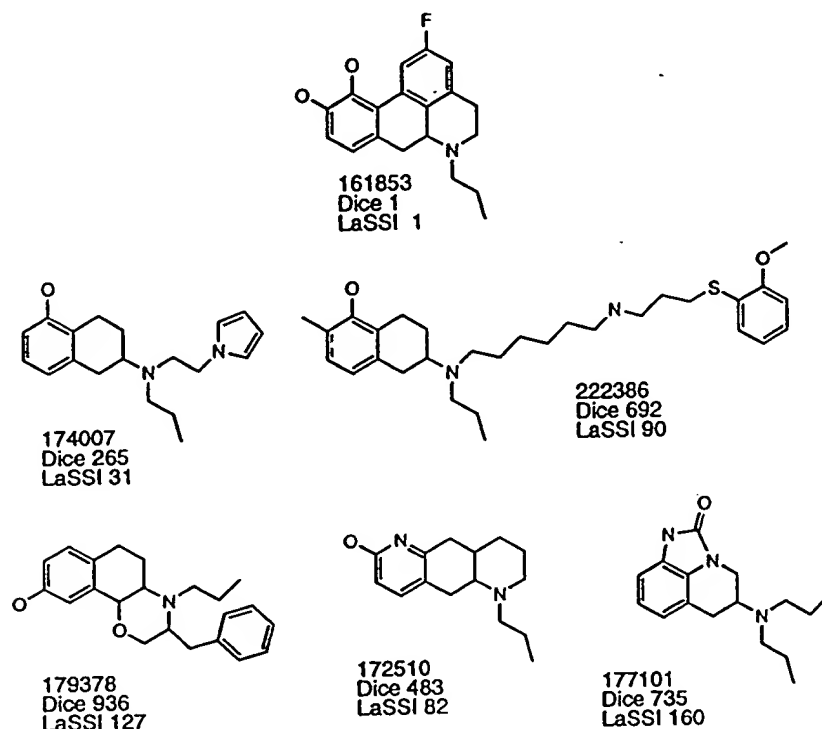
Figure II

Mean similarity of the probe to each molecule in the top scoring 300 compounds (MSP300) for three examples. The MSP300 for Dice searches are shown as a horizontal lines. For comparison, the MSP300 for random sets of 300 compounds from MDDR would be 0.12-0.14.

MSP300

Dice TT

Legend:
- 161853 Dice APTT
- 170534 Dice APTT
- 199183 Dice APTT
- 161853 LaSSI APTT
- 170534 LaSSI APTT
- 199183 LaSSI APTT

Y-axis: Mean similarity to probe for top 300
X-axis: Number of singular values
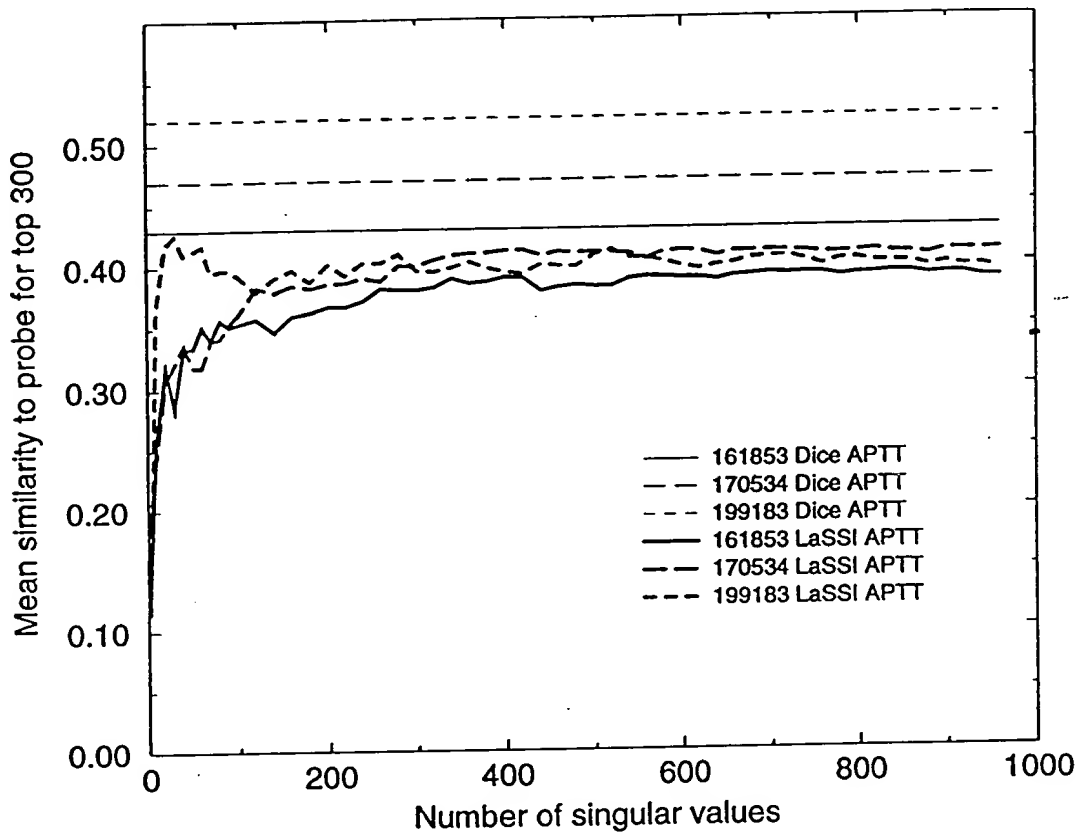
Figure 12

Cumulative actives found vs compounds tested for 187236 as a probe. The actives are oxytocin antagonists that do not contain a dipeptide moiety.

187236 as probe

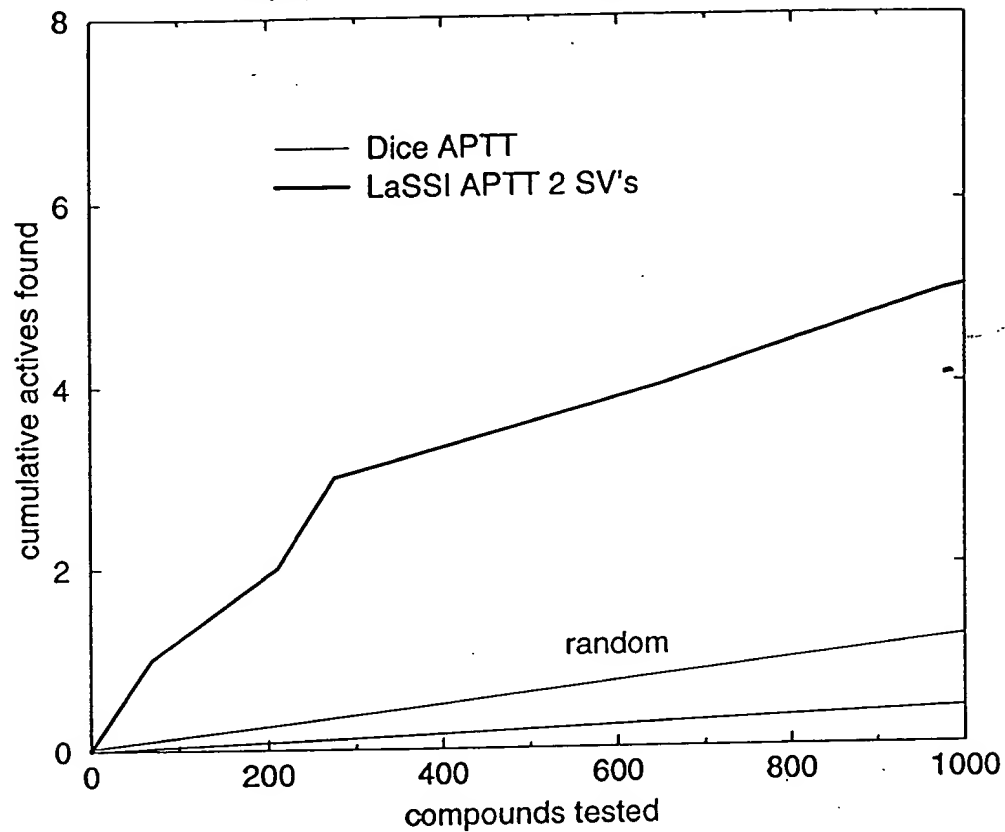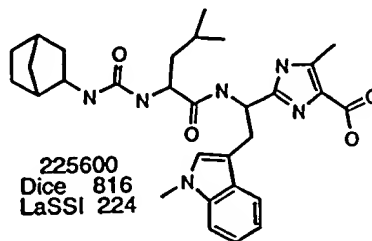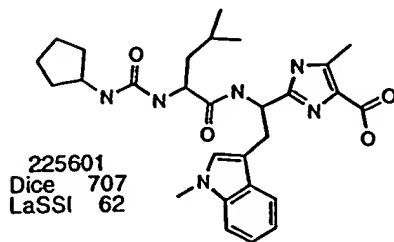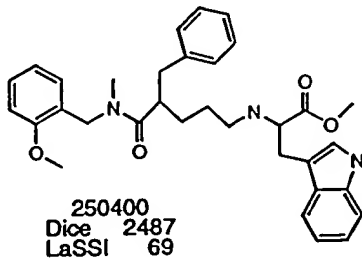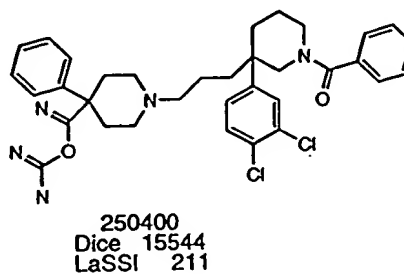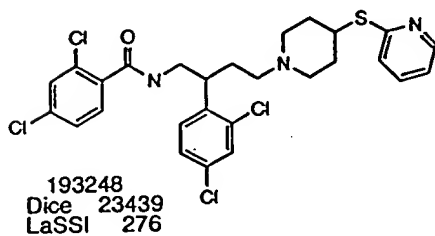### nonpeptide neurokinin antagonists as actives

Figure 13

Selected non-peptide compounds that have extremely different ranks in Dice APTT vs LaSSI APTT for the statistically significant peptide t non-peptide examples.
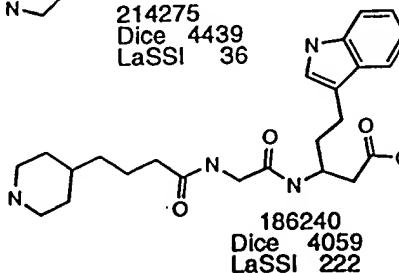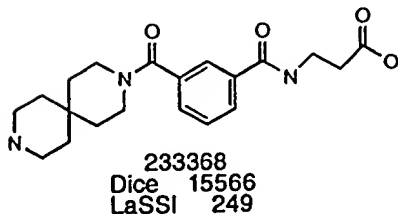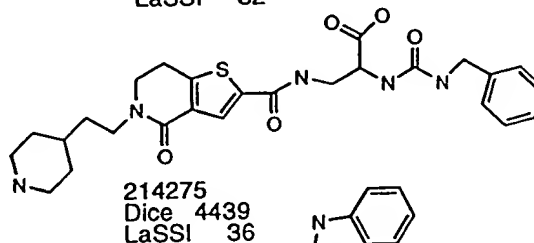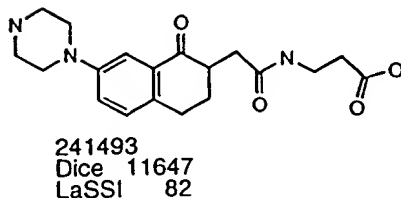
174556 endothelin antagonists (9 SV's)

225601
Dice   707
LaSSI   62

225600
Dice   816
LaSSI   224

187236 neurokinin antagonists (2 SV's)

193248
Dice   23439
LaSSI   276

250400
Dice   15544
LaSSI   211

250400
Dice   2487
LaSSI   69

188541 gpIIb/IIIa receptor antagonists (15 SV's)

211643
Dice   17897
LaSSI   8

199234
Dice   13086
LaSSI   32

241493
Dice   11647
LaSSI   82

214275
Dice   4439
LaSSI   36

233368
Dice   15566
LaSSI   249

186240
Dice   4059
LaSSI   222

FIG.14

FIG. 15

932

FIG. 16